

---

# Protenix-Dock: An accurate and trainable end-to-end protein-ligand docking framework using empirical scoring functions

---

ByteDance AML AI4Science Team

## Abstract

Molecular docking is a fundamental tool for structure-based drug design, enabling the prediction of protein-ligand binding poses and interactions. While classical docking methods offer physically interpretable results, open-source implementations often suffer from limited performance. Deep learning approaches, on the other hand, face challenges with generalization ability. To address these issues, we introduce **Protenix-Dock**, an open-source, trainable, end-to-end classical docking framework built upon a Vina-like Monte Carlo sampling approach and carefully designed scoring functions that integrate both physics-based and empirical terms. Achieving a success rate of 78.9% (top-1 pose RMSD < 2 Å) on PoseBusters V2 dataset, Protenix-Dock demonstrates robust docking performance across diverse targets and ligands. Case studies further demonstrate the effectiveness of its scoring and sampling strategies, and its better generalization ability compared to modern deep learning approaches on unseen binding modes. These findings highlight the potential of Protenix-Dock as a valuable tool in the field of computational drug discovery.

## 1 Introduction

Molecular docking is a computational approach that predicts the binding pose and affinity of a ligand to its target protein. This method provides valuable insights into the structural mechanisms governing protein-ligand interactions, facilitating a deeper understanding of how drugs can be engineered to modulate protein function. As a practical and versatile tool, molecular docking has been extensively employed in virtual screening to identify potential lead compounds, optimize lead molecules, and support various other stages of structure-based drug design (SBDD) [1–4].

Specifically, classical molecular docking first searches for possible binding poses [5, 6]. This process involves a systematic sampling of different conformations and orientations of molecules (we focus on small molecule ligand in this work) within the binding site (pocket). Once the candidate poses are collected, the ranking scoring functions are then applied to prioritize the optimal binding pose. Classical scoring functions consist of terms reflecting the fundamental intermolecular interactions, such as electrostatic, hydrophobic, and hydrogen bonding etc., that occur between a ligand and its receptor at a given binding pose [7, 8]. These terms, conforming the collective knowledge and intuition of medicinal chemists, are often combined in a linear fashion to yield a final binding score. Following the strategy above, a variety of docking packages have been developed over the years, including AutoDock-Vina (Vina) [9], DOCK [10, 11], Glide [12], GOLD [13], many of which are commercially licensed software.

Deep learning approaches, as a promising alternative to classical methods, have broadened the predictive boundaries of molecular docking by leveraging large-scale data and sophisticated architectures [14–19]. Regressive models facilitate more accurate pose and ligand ranking, while generative models enable novel pose predictions and conformational sampling. State-of-the-art deep learning models show a fairly lower root mean squared deviation (RMSD) of docking pose with respect to

experimentally resolved structure, especially when search space is large (e.g., too many rotatable bonds), and have been applied to study, for example, the binding mechanism of serine rich endogenous peptides to a leucine-rich repeat receptor kinase [20]. Despite the high prediction accuracy for machine learning score functions and complex structure prediction methods, the physical plausibility of predicted poses and generalization ability are yet questionable, compared to classical methods. For example, a recent study demonstrated that after removing a key binding interaction, the model still predicted nearly the same ligand pose [21], suggesting that current deep learning methods do not truly capture the underlying physical principles of protein-ligand binding [21, 22]. Moreover, some deep learning methods show diminished performance as ligand similarity between training set and test set decreases, implying potential overfitting[23]. Therefore, a highly accurate, physically explainable classical model is still desirable.

In this work, we present an open-source, trainable, end-to-end classical docking framework **Protenix-Dock** as a complement part of our previous deep learning complex structure prediction model Protenix [24]. Protenix-Dock features a Vina-like Monte Carlo sampling algorithm that systematically explores the conformational space. Our approach employs both physics-based and empirical scoring function terms, whose parameters and coefficients are optimized for two core tasks: docking pose prediction and virtual screening. Our assessments indicate that Protenix-Dock consistently yields robust and reliable docking pose predictions across diverse targets and ligands. These findings demonstrate the potential of Protenix-Dock as a versatile and effective tool for structure-based drug discovery and related structural biology applications.

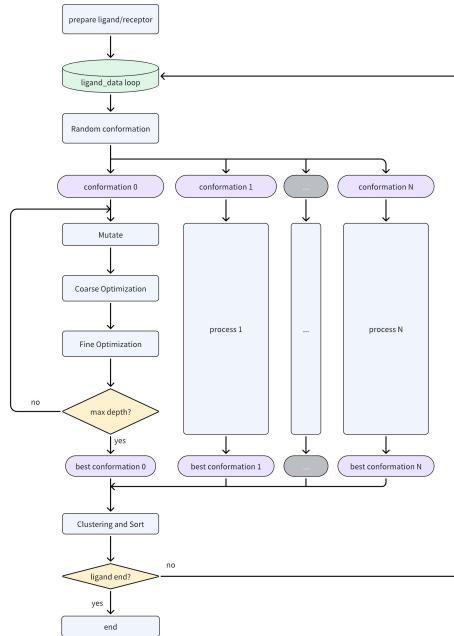
## 2 Method

### 2.1 Docking Pose Sampling

The search of optimal binding pose begins with random ligand conformer initialization using RDKit package[25]. The generated ligand molecules are placed in a predefined docking box with a random position and orientation. The previous step is repeated until a preliminary clash check for the candidate pose is passed. Subsequently, a coarse geometry optimization is performed. If the optimized pose is outside the docking box, indicating that the ligand is in an energetically unfavorable pose, restart from scratch.

The candidate pose is then fed into a Vina-like Monte Carlo (MC) global optimization procedure[9]. The key differences compared to Vina are emphasized as follows:

- We develop an all-atom empirical scoring function for the calculation of metropolis criterion and pose ranking (pose selection score). Meanwhile, in order to ensure the physical plausibility of the sampled poses, we use the total energy derived from the molecular mechanics force field as the objective function in local geometry optimization. The details of the scoring function and force field are described later in section 2.2.
- In each MC step, the mutational degree of freedom is restricted to three kinds of modes, i.e. torsion angle of ligand rotatable bonds, translation and orientation of ligand. On the other hand, during geometry optimization, all cartesian coordinates of ligand atoms and the torsion angle involving certain hydrogen atoms in protein pocket are allowed to relax.
- Local geometry optimizations is performed in two distinct stages. During the mutation



**Figure 1: Monte Carlo Sampling.** Schematic illustration of the workflow for the Protenix-Dock Monte Carlo sampling strategy.

stage, a smaller maximum number of steps is applied to the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [26], allowing rapid energy descent from a potentially distant initial conformation. Once the proposed mutation is accepted, a fine optimization followed, allowing a larger maximum number of BFGS steps to more thoroughly refine the geometry of the pose.

Therefore, the final candidate poses are collected from multiple MC threads, each of which independently executes the described sampling algorithm. Subsequently, these poses are clustered and merged based on an RMSD criterion, similar to the approach developed in Vina, effectively grouping nearly duplicate solutions while retaining structural diversity. The whole MC sampling workflow is shown in Fig. 1.

Alternatively, our package also provides an option that utilizes Vina as a structure sampler. In this workflow, Vina generates initial binding poses, which are subjected to a subsequent geometry optimization step to refine their conformations. After optimization, the poses are scored and ranked using our pose selection scoring functions to identify the most favorable binding configurations.

## 2.2 Scoring Functions

In this work, the scoring functions are formulated as a weighted sum of distinct components that capture various aspects of molecular interactions. Specifically, we combine terms derived from classical force field with contributions from established empirical scoring functions such as Chemscore [27], AA-Score [28]. For a detailed explanation of the mathematical formulation and parameterization of each empirical term, please refer to appendix A and the original publications.

**Force Field** The AMBER FF14SB force field [29] and GAFF2 (version 2.20) [30] are used to describe the intramolecular and intermolecular interactions between proteins and ligands. The force field parameters for HEM cofactors are obtained from [31]. The partial charges of ligand atoms are obtained from AM1-BCC [32, 33] calculations. To evaluate non-bonded interactions accurately while maintaining computational efficiency, Coulomb and van der Waals (vdW) interactions are precomputed on a regular grid encompassing the binding site, and energy values at arbitrary ligand positions are calculated via interpolation from the grid points.

**Interaction vdW energy (with cutoff)** A pairwise vdW cutoff is introduced to accommodate ligand flexibility and reduce the potential energy imposed by harsh steric clashes. Our approach mimics the scaled vdW radii as used in Glide [12], allowing slightly more permissive close contacts.

**Hydrogen bond** Hydrogen bond interactions are modeled using a piecewise linear potential that depends on both the distance between the donor and acceptor atoms and the angle of hydrogen bond [27]. In our implementation, neutral and charged donor–acceptor pairs are treated independently with different parameter sets.

**Hydrophobic interaction** Hydrophobic interactions are described using a piecewise linear potential based on the distance between hydrophobic contacts [27]. Two hydrophobic atoms are considered in contact when the distance between them is within the threshold defined using the  $\sigma$  parameter from the Lennard-Jones (LJ) potential.

**Rotatable Bond Penalty** To penalize excessive flexibility in ligands, the rotatable bond penalty is implemented as equation 18. This penalty increases with the number of rotatable bonds in the ligand, reflecting the entropic cost of binding. By discouraging overly flexible ligands, this term aids in selecting compounds that are more likely to adopt well-defined, energetically favorable conformation, thereby enhancing binding specificity and affinity.

**Desolvation Energy** Our scoring function incorporates a Generalized Born Surface Area (GBSA) energy term to account for desolvation effects accompanying ligand binding [34]. Specifically, we follow the OpenMM [35] implementation of the GBSA energy, in which the electrostatic interaction between solute and solvent is described by the OBC2 model [36] and the non-polar solvation is accounted for by an approximated term of surface area [35, 37]. This GBSA term also partially compensates the overestimated Coulomb interaction for improperly charged ligands.

**$\pi$  Stacking**  $\pi$  stacking interactions are modeled using a piecewise linear potential to capture two key geometries: face-to-face and edge-to-face stacking. The scoring function incorporates essential parameters, such as the center-to-center distance between the aromatic rings, the angle between their planes, and any lateral offset. To account for distinct structural and electronic characteristics, different parameter sets are applied for five- and six-atom aromatic rings [28, 38].

**$\pi$ -Cationic interaction**  $\pi$ -cationic interactions are modeled using a piecewise linear potential, with the key parameter being the distance between the cationic center and the center of the aromatic ring. Similar to  $\pi$  stacking interactions, separate parameter sets are applied for five-atom and six-atom aromatic rings [28, 38].

**Ionic interaction** Ionic interactions between oppositely charged species are additionally modeled using a piecewise linear potential, with the key parameter being the distance between the cation and anion. This term accounts for the required corrections in fixed-charge models for predicting hydration free energy [39].

**Torsion Strain Penalty** Torsion strain penalties are modeled using periodical torsional potential form commonly used in molecular mechanics. This term accounts for the unfavorable energetic efforts ligands have to make to better fit into protein’s binding pockets. The torsion SMARTS patterns are carefully selected from the OpenFF proper torsion patterns [40], which are characterized by local minima conformations that are nearly planar.

We employed three distinct scoring functions, each tailored to specific tasks within the docking workflow (Appendix B). For geometry optimization, the scoring function integrated both intramolecular and intermolecular force field terms to accurately refine the conformations of the ligand. Pose selection utilized a scoring function that combined intermolecular Coulomb and vdW interactions (with cutoff), supplemented by pose-dependent empirical terms to effectively evaluate candidate binding poses. Finally, binding affinity ranking was achieved using a comprehensive scoring function that included all terms mentioned above.

## 2.3 Training of Scoring Functions

### 2.3.1 Dataset

The pdb entries used in our model training were taken from PDBbind (version 2020) [41], DUD-E [42], and the PoseBusters dataset [23]. Instead of using the structures provided by the dataset authors, we redo the preprocessing starting from the original experimental structures curated in the Protein Data Bank (PDB) [43]. The data preparation pipelines are as follows:

- The protonation states of the protein and ligand are determined using the pdb2pqr [44, 45] and OpenBabel packages [46].
- Metal-chelating residues, such as histidine, are kept in their original protonation states. If pdb2pqr alters these states, they are reverted to their original states and hydrogens are added again with pdb2pqr.
- Cofactors (heavy-atom distance  $> 5 \text{ \AA}$ ) and chains (heavy-atom distance  $> 8 \text{ \AA}$ ) far from the ligand of interest are removed.
- Special treatment is applied for non-standard residues and cofactors near the ligand of interest that are not compatible with the AMBER force field: non-standard residues are mutated to alanine; cofactors are removed.
- Crystallization aids and free ions are removed. The definition for crystallization aids can be found in [15].
- Local geometry optimization of the protein-ligand complex is performed using the AMBER force field with OpenMM [35]. The optimization will terminate when the all-atom RMSD exceeds  $0.3 \text{ \AA}$ .
- After the geometry optimization, water molecules are removed.

For the pose selection task, we trained our model on the PDBbind general set and test it on the PDBbind core set and PoseBusters dataset. Duplicate entries were removed from the training set.

For the screening task, we randomly split the DUD-E dataset in an 80:20 ratio as train and valid set. Prior to training, we enumerated the ligand protonation states and stereoisomer with the provided SMILES representations, using the OpenBabel and RDKit packages. This gave a dataset that is more consistent with realistic screening scenarios.

### 2.3.2 Training

The pose selection scoring function (PScore) and affinity ranking scoring function (BScore) were trained sequentially, as the binding affinity ranking relied on evaluating the top-ranked pose. First, 50 candidate poses were generated using Vina sampling, where "good" poses are defined as those with an RMSD relative to the crystal structure less than 2.0 $\text{\AA}$ , while "bad" poses have RMSD values greater than 2.0 $\text{\AA}$ . The primary training objective for PScore was then to prioritize good poses over bad poses. Once an initial PScore was obtained, it could be further refined iteratively using the MC sampling process described earlier, which generated new good and bad poses for additional training cycles. Afterward, PScore was used to identify the best pose for both active and decoy ligands. The training objective for BScore then became ranking active ligands higher than decoy ligands, while different protonation states and stereoisomers of ligands were aggregated using their predicted BScores. Both tasks could be integrated into one unified training framework. The loss function was defined as

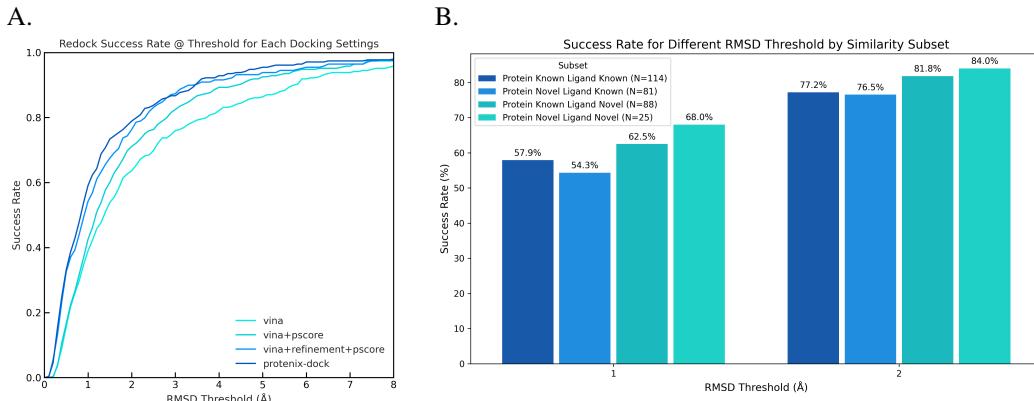
$$Loss = -\log\left(\frac{\exp(SF(\hat{R}_g))}{\exp(SF(\hat{R}_g)) + \sum \exp(SF(\hat{R}_b))}\right)$$

where  $\hat{R}_g$  denotes good pose or active ligand pose,  $\hat{R}_b$  denotes bad pose or decoy ligand pose, respectively. During training procedure, we made all the physical parameters responsible for calculating empirical terms in the scoring functions tunable to enhance the expressive power and adaptability to responding scenarios. The boundary of those parameters were restricted to retain the physical nature of terms. The tunable parameters and the corresponding boundaries in each score terms are listed in Appendix D. We conducted a systematic hyperparameter searching to obtain score functions with best performance. To accelerate the training process, data points whose receptor size is larger than a certain threshold were skipped.

It should be noted that, as we did not use binding affinities as labels, the final scores do not correspond to any practical physical quantity.

## 3 Result

### 3.1 Docking Power



**Figure 2: Redock Success Rate.** **A.** Comparison of protein-ligand redock success rate on Posebusters V2 under different docking settings. "vina+pscore" denotes that the top-1 pose is selected by PScore out of 50 vina generated poses. Similarly, "vina+refinement+pscore" denotes that an additional geometry optimization is performed before PScore ranking. **B.** Success rate for Protenix-Dock evaluated on subset of Posebusters V2 dataset split according to the similarity relative to training dataset. The sequence identity cutoff for protein is 0.4, the Tanimoto similarity (Morgan fingerprint) cutoff for ligand is 0.5, respectively.

We evaluated our method’s docking power by measuring its ability to predict accurate binding poses on the Posebusters Version 2 benchmark set, which comprises 308 carefully selected high-quality protein-ligand complexes from the PDB. The success rate (SR), defined as the percentage of ligands for which the best-scored redock pose has an RMSD below a predetermined threshold, serves as our primary metric. The threshold chosen for the evaluation are 1 Å, which tests the ability to reproduce highly accurate binding poses, and 2 Å, a widely accepted criterion for near-native predictions. The redock RMSD was calculated with respect to the crystal structure.

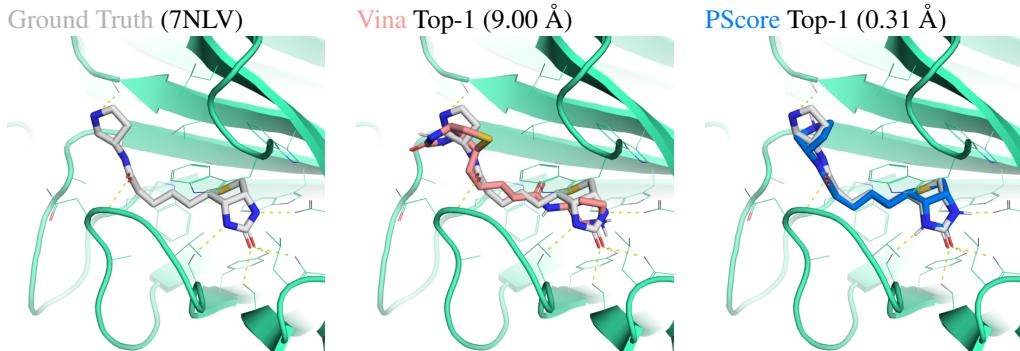
Fig. 2A summarizes the comparative performance of various docking settings. Our method significantly outperforms Vina’s original docking results, delivering improvements of approximately 15% in the SR@1 Å metric and 20% in the SR@2 Å metric, respectively, demonstrating the enhanced ability to predict near-native ligand poses. To dissect the contributions of the sampling algorithm, geometry optimization and the scoring function, we also evaluated Vina-generated poses after two post-processing steps: structural refinement by force field and subsequent rescoring with our PScore model. The results reveal that rescoring Vina poses with PScore alone yields a substantial improvement in performance, boosting the SR@2 Å from 63.64% to 71.10%. Further structural refinement of these poses leads to an additional gain, from 71.10% to 76.30%. Finally, When our MC sampling method is applied, an extra 2.6% improvement is observed. These finding highlight the effectiveness of each component of our approach. Additionally, we examined the success rate on the subsets of the PoseBusters V2 dataset, partitioned based on the similarity to the training set, shown in Fig. 2B. Our method performs even better on novel protein and novel ligand, implying the strong generalization to unseen binding modes, likely due to the physics-informed characteristic of the scoring function.

### 3.2 Screening Power

The results of the BScore evaluation will be released once they are available.

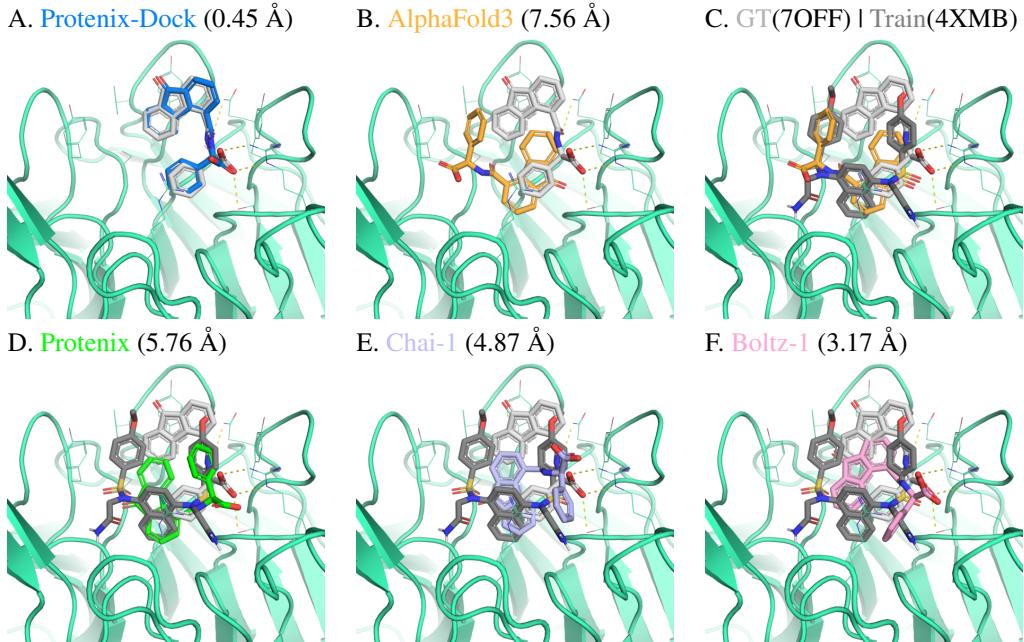
### 3.3 Case Study

Beyond quantitative metrics, we conducted case studies to validate Protein-Dock’s performance improvements on pose scoring and pose sampling. For pose scoring, in 89/308 PoseBusters cases where Vina failed to identify correct poses despite successful sampling (Minimum RMSD < 2 Å in top 50 poses), PScore rescoring recovered 47 cases, while refinement+PScore and MC sampling+PScore both achieved 60 recoveries. Representative cases demonstrate PScore’s improved recognition of hydrogen-bonding networks (Fig. 3), Coulomb interactions (Appendix E Fig. 5A), and π-stacking geometries (Appendix E Fig. 5B). For pose sampling, advanced sampling capability allows successful predictions in 9/23 cases where Vina completely failed to sample near-native poses (RMSD < 2 Å), as shown in Appendix E Fig. 6.



**Figure 3: Evaluation of Pose Scoring Capability.** Wild-type core-streptavidin bound to a conjugated biotinylated pyrrolidine II. Left column: Ground truth structure with protein shown in green-cyan and ligand in light-grey. Middle column: Top-ranked pose selected by Vina scoring function. Right column: After the structure refinement of Vina’s top 50 sampled poses, the top-ranked pose is selected by the PScore scoring function. Heavy-atom RMSD values (Å) are indicated in parentheses. Two more cases in Appendix E Fig. 5

Furthermore, we compared Protenix-Dock with AlphaFold3-like models, focusing on cases where deep learning approaches failed. While direct comparisons were limited (e.g., 7WUY with two ligands in pocket), we identify an interesting case, as shown in Fig. 4. For a novel inhibitor targeting Keap1-Nrf2 (PDB 7OFF [47]) with a binding mode not seen in training data, Protenix-Dock predicted the correct structure with high precision (RMSD = 0.45 Å). In contrast, AlphaFold3, Protenix, and Chai-1 [48] produced poses similar to known training examples like PDB 4XMB; Boltz-1 [49] captured some key interactions but had large errors (RMSD = 3.17 Å). This shows Protenix-Dock’s physics-informed empirical scoring function handles novel protein-ligand binding modes better than deep learning approaches relying heavily on existing data patterns.



**Figure 4: Comparative Analysis of Protenix-Dock and AlphaFold3-like Models.** **A.** Protenix-Dock successfully predicts the binding pose (marble-blue) of a small-molecule inhibitor targeting the Keap1-Nrf2 protein-protein interaction in the Keap1 kelch domain (PDB 7OFF; ground truth shown in light-gray). **B.** AlphaFold3-predicted ligand binding pose for the 7OFF complex from original publication. **C.** Structural comparison between AlphaFold3’s prediction and a training-set ligand conformation (PDB 4XMB, dark-gray). **D-F.** Lowest heavy-atom RMSD poses selected from 25 conformations generated across 5 independent seeds (5 samples per seed) for **D.** Protenix, **E.** Chai-1, and **F.** Boltz-1 models.

## 4 Discussion

In the Protenix-Dock, we introduced a scoring function that incorporates a number of empirical terms with clear physical significance to evaluate the binding affinity between a protein structure and a ligand conformation. As previously detailed in Section 2.2, extensive work has been done to explain the importance of each component in the scoring function. Therein, to further validate the necessity of each component, we conducted ablation studies by systematically removing individual components of the scoring function, as shown in table 1.

**Table 1: Ablation Study on different Empirical Terms.** The success rate of top-1 pose RMSD < 2 Å is calculated for different term settings. Default settings are listed on the left, +gbsa means adding GBSA terms in the scoring functions, and columns like -vdW means removing responding terms.

Dataset	Baseline	Terms Added or Removed During Training								
		+ gbsa	- vdW	- hbond	- coul	- lipo	- torsion	- pi-cation	- ionic	- pi-stack
PB-v2	76.30	75.32	53.25	73.70	75.65	76.30	75.97	76.30	76.30	76.30
Pdbbind Core	70.46	71.89	51.60	67.97	69.40	68.33	70.46	69.75	70.46	70.11

The results indicate that van der Waals (vdW) and hydrogen-bond (hbond) interactions are crucial for successful binding, as removing either leads to a significant drop in the success rate. In contrast, components such as Coulomb, hydrophobic (lipo) interactions, and torsion strain penalties are found to be less important, which affect the overall success rate limitedly. The least significant components are identified as  $\pi$ -stacking,  $\pi$ -cationic, presumably because of the lower occurrence frequency of these interactions [8], and ionic interactions, which might already be covered by coulomb interaction. Notably, the addition of GBSA terms did not enhance performance in the PoseBusters V2 benchmark set and showed limited improvement in PDBBind Core set. This result demonstrates the robustness of our PScore and supports the rationale of removing GBSA terms in PScore.

## 5 Limitation and Future Work

Rigid docking approaches come with inherent limitations that can affect their applicability and accuracy. One primary constraint is the reliance on a resolved receptor structure, as a fixed conformation of the binding pocket is assumed. This can be problematic for proteins that exhibit significant flexibility or adopt alternative conformations upon ligand binding. Additionally, ligands with flexible ring systems or large conformational spaces pose challenges, since classical rigid docking cannot efficiently sample all possible conformations, potentially overlooking favorable binding poses.

In the case of Protenix-Dock, specific limitations arise from its design and implementation. The inclusion of GBSA terms for desolvation effects, while enhancing accuracy, introduces high computational complexity, making it less efficient for large-scale docking tasks. Another challenge lies in its dependency on predefined force fields, which may lack parameters for unusual residue or molecule types, reducing its versatility for non-standard systems. Moreover, Protenix-Dock cannot handle ligands that form covalent bonds with the receptor, limiting its utility in studying irreversible inhibitors or covalent drug candidates. Finally, the local minima predicted by the force field during energy minimization may differ from experimentally resolved crystal structures, potentially leading to deviations in ligand-receptor geometry and reducing predictive accuracy. These factors highlight the trade-offs inherent in the use of Protenix-Dock in general.

To overcome the limitations, our future plan would mainly focus on the integration of physics-based models and deep learning. We expect to create a synergistic system that combines the interpretability of physics-based methods with the adaptability and predictive power of machine learning, paving the way for more reliable and scalable tools in drug discovery.

## 6 Code Availability

Source code is available at the Protenix-Dock repository (<https://github.com/bytedance/Protenix-Dock>)

## 7 Contribution

Author List (Alphabetical Order): Xinshi Chen, Zhenyu Liu, Bo Shi, Shaochen Shi, Wenzhi Xiao, Jiabo Xu, Xingyuan Xu, Jincai Yang, Yuxuan Zhang, Ziqing Zhang.

W. Xiao, Z. Liu, X. Chen proposed the concept of the work; J. Yang, X. Xu acquired and prepared the raw data; Z. Liu, J. Xu, X. Xu, X. Chen, Y. Zhang designed the scoring functions; Z. Zhang, X. Chen designed the sampling algorithm; X. Chen, Y. Zhang built the training code. J. Xu, Z. Liu performed the fitting; B. Shi, S. Shi created the c++ engine; B. Shi, S. Shi, Z. Liu prepared the github repository; Z. Liu, J. Xu, J. Yang analyzed the result; Z. Liu, J. Xu, J. Yang, X. Xu, B. Shi drafted the manuscript.

## 8 Acknowledgment

We acknowledge the help from Wenzhi Ma who provided inference results of Chai-1, Boltz-1 and Protenix models. We also acknowledge the help from Jiahui Tong and Youjun Xu who provided thoughtful discussion about the classical docking method and the biochemical insights of the cases.

## References

- [1] Michael K Gilson and Huan-Xiang Zhou. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.*, 36(1):21–42, 2007.
- [2] Jiankun Lyu, Sheng Wang, Trent E Balius, Isha Singh, Anat Levit, Yurii S Moroz, Matthew J O’Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, et al. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224–229, 2019.
- [3] Jiankun Lyu, Nicholas Kapolka, Ryan Gumpfer, Assaf Alon, Liang Wang, Manish K Jain, Ximena Barros-Álvarez, Kensuke Sakamoto, Yoojoong Kim, Jeffrey DiBerto, et al. AlphaFold2 structures guide prospective ligand discovery. *Science*, page eadn6354, 2024.
- [4] Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, Xiaofeng Li, Leike Zhang, Chao Peng, et al. Structure of mpro from sars-cov-2 and discovery of its inhibitors. *Nature*, 582(7811):289–293, 2020.
- [5] Brian J Bender, Stefan Gahbauer, Andreas Luttens, Jiankun Lyu, Chase M Webb, Reed M Stein, Elissa A Fink, Trent E Balius, Jens Carlsson, John J Irwin, et al. A practical guide to large-scale docking. *Nature protocols*, 16(10):4799–4832, 2021.
- [6] Joseph M Paggi, Ayush Pandit, and Ron O Dror. The art and science of molecular docking. *Annual Review of Biochemistry*, 93, 2024.
- [7] Chao Shen, Junjie Ding, Zhe Wang, Dongsheng Cao, Xiaoqin Ding, and Tingjun Hou. From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(1):e1429, 2020.
- [8] Renato Ferreira de Freitas and Matthieu Schapira. A systematic analysis of atomic protein–ligand interactions in the pdb. *Medchemcomm*, 8(10):1970–1981, 2017.
- [9] Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.
- [10] Ryan G Coleman, Michael Carchia, Teague Sterling, John J Irwin, and Brian K Shoichet. Ligand pose and orientational sampling in molecular docking. *PloS one*, 8(10):e75992, 2013.
- [11] William J Allen, Trent E Balius, Sudipto Mukherjee, Scott R Brozell, Demetri T Moustakas, P Therese Lang, David A Case, Irwin D Kuntz, and Robert C Rizzo. Dock 6: Impact of new features and current docking performance. *Journal of computational chemistry*, 36(15):1132–1156, 2015.
- [12] Richard A Friesner, Jay L Banks, Robert B Murphy, Thomas A Halgren, Jasna J Klicic, Daniel T Mainz, Matthew P Repasky, Eric H Knoll, Mee Shelley, Jason K Perry, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7):1739–1749, 2004.
- [13] Gareth Jones, Peter Willett, Robert C Glen, Andrew R Leach, and Robin Taylor. Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, 267(3):727–748, 1997.
- [14] Rohith Krishna, Jue Wang, Woody Ahern, Pascal Sturmels, Preetham Venkatesh, Indrek Kalvet, Gyu Rie Lee, Felix S Morey-Burrows, Ivan Anishchenko, Ian R Humphreys, et al. Generalized biomolecular modeling and design with rosettafold all-atom. *Science*, 384(6693):eadl2528, 2024.
- [15] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphaFold 3. *Nature*, pages 1–3, 2024.
- [16] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

- [17] Eric Alcaide, Zhifeng Gao, Guolin Ke, Yaqi Li, Linfeng Zhang, Hang Zheng, and Gengmo Zhou. Uni-mol docking v2: Towards realistic and accurate binding pose prediction. *arXiv preprint arXiv:2405.11769*, 2024.
- [18] Hongjian Li, Kam-Heung Sze, Gang Lu, and Pedro J Ballester. Machine-learning scoring functions for structure-based drug lead optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(5):e1465, 2020.
- [19] Hongjian Li, Kam-Heung Sze, Gang Lu, and Pedro J Ballester. Machine-learning scoring functions for structure-based virtual screening. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(1):e1478, 2021.
- [20] Simon Snoeck, Hyun Kyung Lee, Marc W Schmid, Kyle W Bender, Matthias J Neeracher, Alvaro D Fernández-Fernández, Julia Santiago, and Cyril Zipfel. Leveraging coevolutionary insights and ai-based structural modeling to unravel receptor-peptide ligand-binding mechanisms. *Proceedings of the National Academy of Sciences*, 121(33):e2400862121, 2024.
- [21] Matthew R. Masters, Amr H. Mahmoud, and Markus A. Lill. Do deep learning models for co-folding learn the physics of protein-ligand interactions? *bioRxiv*, 2024. doi: 10.1101/2024.06.03.597219.
- [22] David Errington, Constantin Schneider, Cédric Bouysset, and Frédéric A. Dreyer. Assessing interaction recovery of predicted protein-ligand poses, 2024. URL <https://arxiv.org/abs/2409.20227>.
- [23] Martin Buttenschoen, Garrett M Morris, and Charlotte M Deane. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 2024.
- [24] Xinshi Chen, Yuxuan Zhang, Chan Lu, Wenzhi Ma, Jiaqi Guan, Chengyue Gong, Jincai Yang, Hanyu Zhang, Ke Zhang, Shenghao Wu, Kuangqi Zhou, Yanping Yang, Zhenyu Liu, Lan Wang, Bo Shi, Shaochen Shi, and Wenzhi Xiao. Proteinix - advancing structure prediction through a comprehensive alphafold3 reproduction. *bioRxiv*, 2025. doi: 10.1101/2025.01.08.631967.
- [25] Rdkit: Open-source cheminformatics. URL <http://www.rdkit.org>.
- [26] Stephen J. Wright Jorge Nocedal. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006. doi: 10.1007/978-0-387-40065-5.
- [27] Matthew D Eldridge, Christopher W Murray, Timothy R Auton, Gaia V Paolini, and Roger P Mee. Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of computer-aided molecular design*, 11:425–445, 1997.
- [28] Xiaolin Pan, Hao Wang, Yueqing Zhang, Xingyu Wang, Cuiyu Li, Changge Ji, and John ZH Zhang. Aa-score: a new scoring function based on amino acid-specific interaction for molecular docking. *Journal of Chemical Information and Modeling*, 62(10):2499–2509, 2022.
- [29] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.
- [30] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.
- [31] Debra Ann Giammona. *An examination of conformational flexibility in porphyrins and bulky-ligand binding in myoglobin*. University of California, Davis, 1984.
- [32] Araz Jakalian, Bruce L Bush, David B Jack, and Christopher I Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: I. method. *Journal of computational chemistry*, 21(2):132–146, 2000.

- [33] Araz Jakalian, David B Jack, and Christopher I Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: II. parameterization and validation. *Journal of computational chemistry*, 23(16):1623–1641, 2002.
- [34] Tingjun Hou, Junmei Wang, Youyong Li, and Wei Wang. Assessing the performance of the mm/pbsa and mm/gbsa methods. I. the accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling*, 51(1):69–82, 2011.
- [35] Peter Eastman, Raimondas Galvelis, Raúl P Peláez, Charles RA Abreu, Stephen E Farr, Emilio Gallicchio, Anton Gorenko, Michael M Henry, Frank Hu, Jing Huang, et al. Openmm 8: molecular dynamics simulation with machine learning potentials. *The Journal of Physical Chemistry B*, 128(1):109–116, 2023.
- [36] Alexey Onufriev, Donald Bashford, and David A Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics*, 55(2):383–394, 2004.
- [37] Michael Schaefer, Christian Bartels, and Martin Karplus. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *Journal of molecular biology*, 284(3):835–848, 1998.
- [38] Sebastian Salentin, Sven Schreiber, V Joachim Haupt, Melissa F Adasme, and Michael Schroeder. Plip: fully automated protein–ligand interaction profiler. *Nucleic acids research*, 43(W1):W443–W447, 2015.
- [39] David L Mobley, Elise Dumont, John D Chodera, and Ken A Dill. Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *The Journal of Physical Chemistry B*, 111(9):2242–2254, 2007.
- [40] Simon Boothroyd, Pavan Kumar Behara, Owen C Madin, David F Hahn, Hyesu Jang, Vytautas Gapsys, Jeffrey R Wagner, Joshua T Horton, David L Dotson, Matthew W Thompson, et al. Development and benchmarking of open force field 2.0. 0: the sage small molecule force field. *Journal of chemical theory and computation*, 19(11):3251–3275, 2023.
- [41] Renxiao Wang, Xueliang Fang, Yipin Lu, and Shaomeng Wang. The pdbind database: Collection of binding affinities for protein- ligand complexes with known three-dimensional structures. *Journal of medicinal chemistry*, 47(12):2977–2980, 2004.
- [42] Michael M Mysinger, Michael Carchia, John J Irwin, and Brian K Shoichet. Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, 55(14):6582–6594, 2012.
- [43] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.
- [44] Todd J Dolinsky, Jens E Nielsen, J Andrew McCammon, and Nathan A Baker. Pdb2pqr: an automated pipeline for the setup of poisson–boltzmann electrostatics calculations. *Nucleic acids research*, 32:W665–W667, 2004.
- [45] Todd J Dolinsky, Paul Czodrowski, Hui Li, Jens E Nielsen, Jan H Jensen, Gerhard Klebe, and Nathan A Baker. Pdb2pqr: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research*, 35:W522–W525, 2007.
- [46] Noel M O’Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3:1–14, 2011.
- [47] Dilip Narayanan, Kim T Tran, Jakob S Pallesen, Sara MØ Solbak, Yuting Qin, Elina Mukminova, Martina Luchini, Kristina O Vasilyeva, Dorleta González Chichón, Georgia Goutsiou, et al. Development of noncovalent small-molecule keap1-nrf2 inhibitors by fragment-based drug discovery. *Journal of medicinal chemistry*, 65(21):14481–14526, 2022.

- [48] Chai Discovery, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhnikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *BioRxiv*, pages 2024–10, 2024.
- [49] Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, et al. Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv*, pages 2024–11, 2024.
- [50] Cédric Bouyssset and Sébastien Fiorucci. Prolif: a library to encode molecular interactions as fingerprints. *Journal of cheminformatics*, 13(1):72, 2021.

## A The mathematical formulation of each empirical terms

The formula of empirical terms are listed in Table 2. Many terms involve a distance/angle function  $g(x)$ , which is a piecewise linear function defined as

$$g(x) = \begin{cases} 1 & x < x_1 \\ 1 - (x - x_1)/(x_2 - x_1) & x_1 \leq x < x_2 \\ 0 & x_2 \geq x_2 \end{cases} \quad (1)$$

Table 2: Mathematical formula.

Empirical Terms	Formula	Explanation
Hydrophobic Interaction	$\Delta G_{\text{lipo}} = \sum_{i,j} g(r_{ij}) \quad (2)$ $R_1 = (\sigma_i^{\text{vdw}} + \sigma_j^{\text{vdw}}) \cdot \beta_{\text{lipo}} + R_{\min} \quad (3)$ $R_2 = R_1 + R_{\max} \quad (4)$	The $R_1$ and $R_2$ here in the lower/upper threshold in Eq.(1), and $\sigma^{\text{vdw}}$ is the vdw L-J radius factor of the corresponding atom.
Hydrogen Bond	$\Delta G_{\text{hbond}} = \sum_n^{N_{\text{types}}} c_{\text{hbond}}^n \sum_{i,j} \Delta G_{\text{hbond},n}^{ij} \quad (5)$ $\Delta G_{\text{hbond},n}^{ij} = g_1(\Delta r_{ij}^n) g_2(\Delta \alpha_{ij}^n) \quad (6)$ $\Delta r_{ij}^n =  r_{ij} - R_0^n  \quad (7)$ $\Delta \alpha_{ij}^n =  \alpha_{ij} - \alpha_0^n  \quad (8)$	we classify the hydrogen bond into 8 sub-classes based on the whether the corresponding <b>donor/acceptor atoms are neutral or charged</b> , and whether the donor are from <b>ligand or receptor</b> . Each sub-terms are associated with a coefficient to form the overall hydrogen bond interaction. $R_0$ and $\alpha_0$ are the parameter defining the optimal value of the bonding distance/angle.
Ionic Interactions	$\Delta G_{\text{ionic}} = c_{\text{ion}}^1 \Delta G_{\text{ionic-anion}}^{\text{cation-anion}} + c_{\text{ion}}^2 \Delta G_{\text{ionic}}^{\text{anion-cation}} \quad (9)$ $\Delta G_{\text{ionic}}^{A,B} = \sum_{i_A, j_B} g(r_{ij}) \quad (10)$	We classify the ionic interaction into two sub-classes depending on the <b>location of the cation-anion pairs</b> . Here, the superscript cation-anion indicate the cation is from ligand with anion is from receptor, and vice versa.
$\pi$ -Cationic Interaction	$\Delta G_{\text{pi-cation}} = \sum_n^{N_{\text{types}}} c_{\text{pi-cation}}^n \sum_{ij} g(r_{ij}^n) \quad (11)$ $r_{ij} =  \mathbf{r}_i - \bar{\mathbf{r}}_j  \quad (12)$	We classify the $\pi$ -Cation into 4 sub-classes depending on whether the cation is <b>located on ligand or receptor</b> , and whether the $\pi$ ring is <b>five or six member ring</b> . The $r_{ij}$ distance is the distance between ion and the center of the $\pi$ -ring, which is the average of ring-atom coordinates.
$\pi$ Stacking	$\Delta G_{\text{pi-stack}} = \sum_n^{N_{\text{types}}} c_{\text{pi-stack}}^n \sum_{ij} g(r_{ij}) g(\alpha_{ij}) g(\sigma_{ij}) \quad (13)$ $r_{ij} =  \mathbf{r}_{ij}  =  \mathbf{r}_i - \bar{\mathbf{r}}_j  \quad (14)$ $p_{ij} = \sqrt{r_{ij}^2 - \frac{ \mathbf{r}_{ij} \cdot \mathbf{v}_i ^2}{ \mathbf{v}_i^2 }} \quad \sigma_{ij} = \min(p_{ij}, p_{ji}) \quad (15)$	We classify the $\pi$ -Stacking into 8 sub-classes depending on the stacking type ( <b>Edge-To-Face / Face-To-Face</b> ) and size of the $\pi$ -rings ( <b>five or six member ring</b> ). $\alpha_{ij}$ is the angle between $\mathbf{v}_i$ and $\mathbf{v}_j$ , which are the normal vector of the two $\pi$ -rings, and $\sigma_{ij}$ is projection of $\mathbf{r}_{ij}$ , the distance vector of two $\pi$ -ring centers, on the normal vector $\mathbf{v}_j$ .
Torsion Strain	$\Delta G_{\text{tstrain}} = \sum_n^{N_{\text{patt}}} \sum_{t=1}^{N_t} \sum_{m=1}^4 k_m^n (1 + \cos(m\theta_t^n - \theta_{m,0}^n)) \quad (16)$ $k_m = \kappa_m^n \cdot k_m^{n,0} \quad (17)$	we prepare 19 torsion strain patterns which are collected from OpenFF[40], the patterns are listed in Table 5. We adopted the commonly used proper torsion energy form in empirical force fields. $k_m^n$ is the initial parameter, and $\kappa_m^n$ is the scaling factor. $N_t$ is the number of matched dihedrals for a given torsion pattern.
Rotatable Penalty	$\Delta G_{\text{rotopen}} = \frac{-1}{c + N_{\text{rot}}} \quad (18)$	$N_{\text{rot}}$ is the number of rotatable bonds, and $c$ is a predefined constant.

## B Terms in scoring functions

Table 3: Terms in scoring functions

	Terms	MM Energy	PScore	BScore
Force Field	Bond	✓		
	Angle	✓		
	Torsion	✓		
	Coulomb	✓		
	vdW	✓		
	Interaction Coulomb	✓	✓	✓
Empirical	Interaction vdW	✓	energy cutoff	energy cutoff
	Ionic		✓	✓
	$\pi$ -Stacking		✓	✓
	Hydrogen Bond		✓	✓
	Hydrophobic		✓	✓
	$\pi$ -Cationic		✓	✓
	Torsion Strain Penalty		✓	✓
Desolvation	Rotatable Bond Penalty			✓
	GBSA			✓

## C SMARTS patterns used to define atom types in each empirical term

Table 4: Patterns and SMARTS for empirical terms. SMARTS are derived from ProLIF [50] and RDKit [25].

Empirical Term	Pattern	SMARTS
Hydrophobic Interactions	Hydrophobic Atoms	<chem>[c,s,Br,I,S&amp;H0&amp;v2,\$([D3,D4:#6])&amp;!\$([#6]~[#7,#8,#9])&amp;!\$([#6X4H0]);+0]</chem>
Hydrogen Bond	Donor	<chem>[\$([O,S;+0]),\$([N;v3,v4&amp;+1]),n+0]-[H]</chem>
	Acceptor	<chem>[#7&amp;!\$([nX3])&amp;!\$([NX3]-*[O,N,P,S])&amp;!\$([NX3]-[a])&amp;!\$([Nv4&amp;+1]),o&amp;!\$([OX2](C)C=O)&amp;!\$([O(~a)~a]&amp;!\$([O=N-*)&amp;!\$([O-]-N=O),o+0])</chem>
Ionic Interaction	Anion Atoms	<chem>[-{1-},\$(O=[C,S,P]-[O-])]</chem>
	Cation Atoms	<chem>[+{1-},\$([NX3&amp;!\$([NX3]-O)]-[C]=[NX3+])]</chem>
$\pi$ -Cationic Interaction	Cation Atoms	the same as above
	Five-member $\pi$ -ring	<chem>[a;r5]1:[a;r5]:[a;r5]:[a;r5]:[a;r5]:1</chem>
	Six-member $\pi$ -ring	<chem>[a;r6]1:[a;r6]:[a;r6]:[a;r6]:[a;r6]:[a;r6]:1</chem>
$\pi$ Stacking	Five-member $\pi$ -ring	the same as above
	Six-member $\pi$ -ring	the same as above
Rotatable Bond Penalty	Rotatable Bonds	<chem>[!\$(*##*)&amp;!D1]-&amp;!@[\$(*##*)&amp;!D1]</chem>

Table 5: Torsion strain patterns. SMARTS are derived from OpenFF[40]

Pattern Index	SMARTS
0	[#6X4:1] - [#6X4;r3:2] - [#6X3:3] ~ [#6X3:4]
1	[#6X3:1] - [#6X4;r3:2] - [#6X3:3] = [#8X1:4]
2	[#6X4;r3:1] - ;@[#6X4;r3:2] - [#6X3;r6:3] : [#6X3;r6:4]
3	[#6X4;r3:1] - ;@[#6X4;r3:2] - [#6X3:3] = [#8X1:4]
4	[*:1] ~ [#6X3:2] - [#6X3:3] ~ [*:4]
5	[*:1] ~ [#6X3:2] - [#6X3\$(*=[#8,#16,#7]):3] ~ [*:4]
6	[#6X3:1] = [#6X3:2] - [#6X3:3] = [#8X1:4]
7	[*:1] - [#7X3:2] - [#6X3\$(*=[#8,#16,#7]):3] ~ [*:4]
8	[*:1] - [#7X3:2] - !@[#6X3:3] (= [#8,#16,#7:4]) - [#6,#1]
9	[#1:1] - [#7X3:2] - !@[#6X3:3] (= [#8,#16,#7:4]) - [#6,#1]
10	[*:1] - [#7X3:2] - !@[#6X3:3] (= [#8,#16,#7:4]) - [#7X3]
11	[*:1] = [#7X2,#7X3+1:2] - [#6X3:3] = , : [*:4]
12	[#7X2:1] ~ [#7X2:2] - [#6X3:3] ~ [#6X3:4]
13	[*:1] ~ [#6X3:2] - [#8X2:3] - [*:4]
14	[*:1] ~ [#6X3:2] (= [#8,#16,#7]) - [#8X2H0:3] - [*:4]
15	[*:1] ~ [#6X3:2] (= [#8,#16,#7]) - [#8:3] - [#1:4]
16	[*:1] ~ [#7X2:2] - [#7X3:3] ~ [*:4]
17	[*:1] = [#7X2:2] - [#7X2:3] = [*:4]
18	[*:1] - [#7:2] - [#15:3] = [*:4]

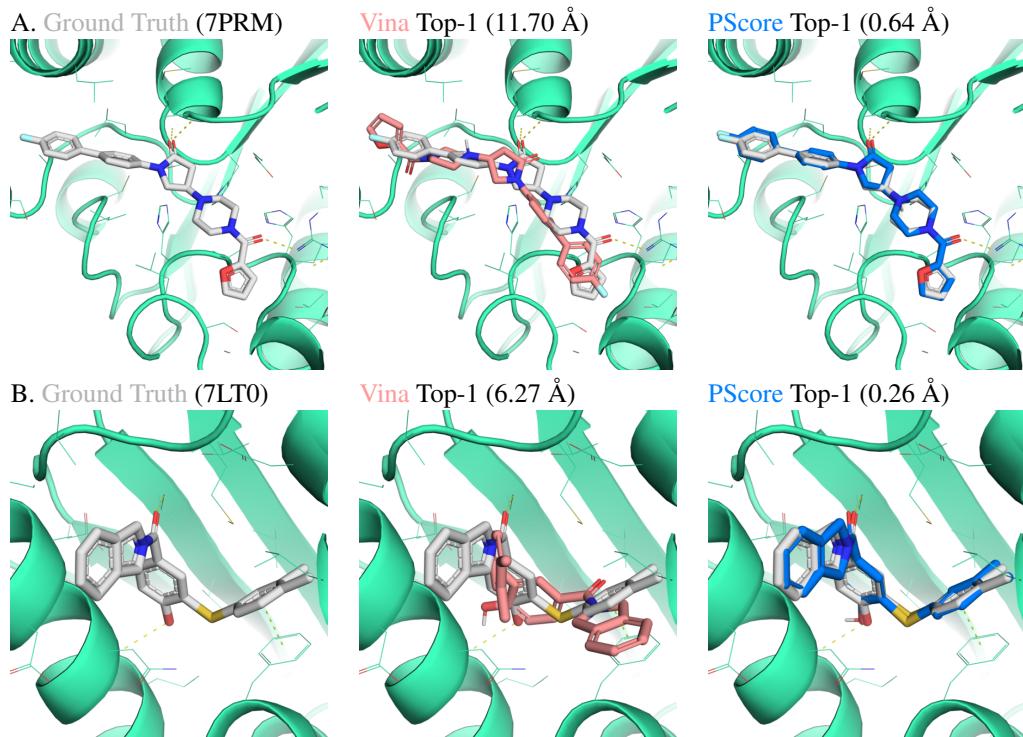
## D Tunable parameters and the corresponding boundaries in each empirical terms

Table 6: Initialization and boundaries of tunable parameters in each empirical terms.

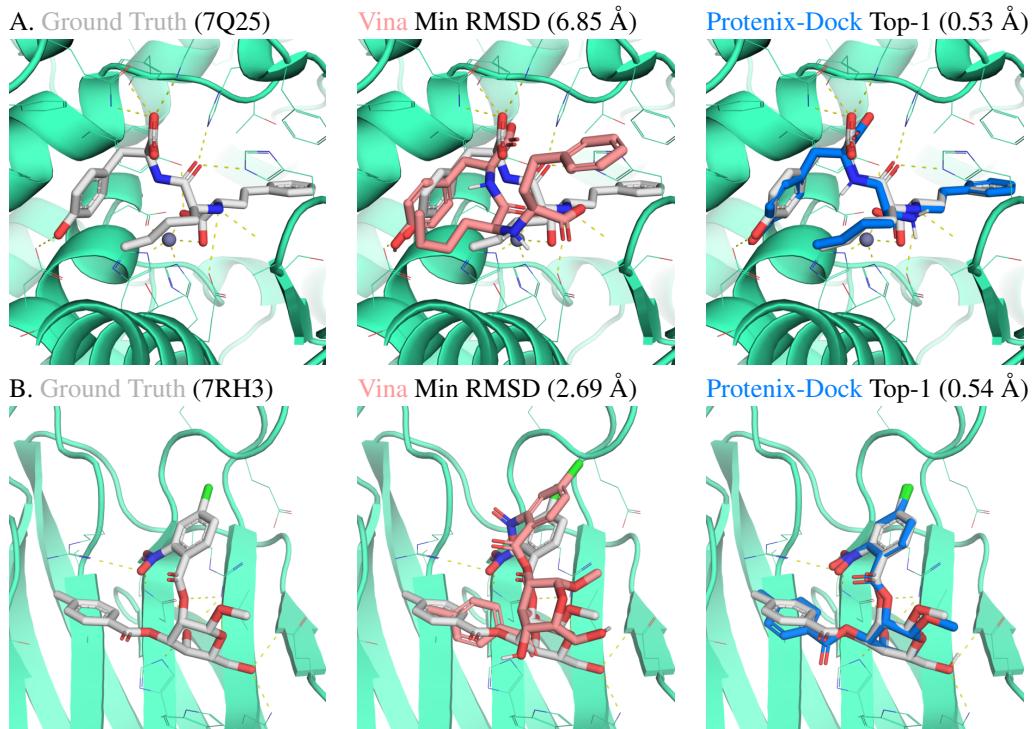
Empirical Term	Parameter	Initial Value	Tunable Range
Hydrophobic Interactions	hydroph_dist_min $R_{\min}$	0.5	[0.2, 0.8]
	hydroph_dist_max $R_{\max}$	2.0	[2.0, 4.0]
	hydroph_sigma_scale $\beta_{lipo}$	1.1	[0.2, 2.0]
Hydrogen Bond	hbond_\${type}\$_coef $c_{\text{hbond}}^n$	1.0	[0.1, 1.0]
	hbond_\${type}\$_dist_min $R_{\text{hbond},1}^n$	0.25	[0.15, 0.35]
	hbond_\${type}\$_dist_max $R_{\text{hbond},2}^n$	0.65	[0.55, 0.75]
	hbond_\${type}\$_dist_center $R_{\text{hbond},0}^n$	1.9	[1.6, 2.2]
	hbond_\${type}\$_angle_min $\alpha_{\text{hbond},1}^n$	30	[20, 40]
	hbond_\${type}\$_angle_max $\alpha_{\text{hbond},2}^n$	80	[70, 90]
Van der Waals	vdw_energy_cutoff $E_c^{\text{vdw}}$	1.0	[0.3, 3.0]
	lig_\${atom_type}_sigma_scale $\eta_\sigma^i$	1.0	[0.8, 1.2]
	lig_\${atom_type}_epsilon_scale $\eta_\epsilon^i$	1.0	[0.5, 1.2]
Ionic Interactions	\${\{ion\_pair\_type}\}\$_coef $c_{\text{ion}}^n$	1.0	[0.1, 1.0]
	\${\{ion\_pair\_type}\}\$_dist_min $R_{\text{ion},1}^n$	3.5	[2.5, 4.5]
	\${\{ion\_pair\_type}\}\$_dist_max $R_{\text{ion},2}^n$	5.5	[4.5, 6.5]
$\pi$ Cation Interactions	\${\{\pi\_cation\_type}\}\$_coef $c_{\text{pi-cation}}^n$	1.0	[0.1, 1.0]
	\${\{\pi\_cation\_type}\}\$_dist_min $R_{\text{pi-cation},1}^n$	3.5	[2.5, 4.5]
	\${\{\pi\_cation\_type}\}\$_dist_max $R_{\text{pi-cation},1}^n$	5.5	[4.5, 6.5]
$\pi$ -Stacking	\${\{\pi\_stacking\_type}\}\$_coef $c_{\text{pi-stack}}^n$	1.0	[0.1, 1.0]
	\${\{\pi\_stacking\_type}\}\$_dist_min $R_{\text{pi-stack},1}^n$	3.5	[2.5, 4.5]
	\${\{\pi\_stacking\_type}\}\$_dist_max $R_{\text{pi-stack},2}^n$	5.5	[4.5, 6.5]
	\${\{\pi\_stacking\_type}\}\$_theta_min $\theta_{\text{pi-stack},1}^n$	20	[15, 25]
	\${\{\pi\_stacking\_type}\}\$_theta_max $\theta_{\text{pi-stack},2}^n$	30	[25, 35]
	\${\{\pi\_stacking\_type}\}\$_sigma_min $\sigma_{\text{pi-stack},1}^n$	2.0	[1.5, 2.5]
	\${\{\pi\_stacking\_type}\}\$_sigma_max $\sigma_{\text{pi-stack},2}^n$	3.0	[2.5, 3.5]
Torsion Strain Penalty	\${\{\text{torsion\_pattern}\}\$_k1_scale $\kappa_m^1$	1.0	[0.8, 1.2]
	\${\{\text{torsion\_pattern}\}\$_k2_scale $\kappa_m^2$	1.0	[0.8, 1.2]
	\${\{\text{torsion\_pattern}\}\$_k3_scale $\kappa_m^3$	1.0	[0.8, 1.2]
	\${\{\text{torsion\_pattern}\}\$_k4_scale $\kappa_m^4$	1.0	[0.8, 1.2]

In the above table, the symbols like \${\{type}\}\$ are variables referring to the subclass of the corresponding interactions. The principle for dividing subclass are mentioned in Appendix A. The patterns of torsion strain penalty are listed in Table 5.

## E Figures of Case Study



**Figure 5: Evaluation of Pose Scoring Capability.** Comparative case study demonstrating Vina’s failure versus PScore’s success in near-native pose identification. Left column: Ground truth structure with protein shown in green-cyan and ligand in light-grey. Middle column: Top-ranked pose selected by Vina scoring function. Right column: After the structure refinement of Vina’s top 50 sampled poses, the top-ranked pose is selected by the PScore scoring function. Heavy-atom RMSD values ( $\text{\AA}$ ) are indicated in parentheses. **A.** Human monoglyceride lipase in complex with compound 13. **B.** N-terminal domain of Hsp90 $\alpha$  complexed with a selective inhibitor.



**Figure 6: Evaluation of Pose Sampling Capability.** Comparative case study demonstrating Vina’s failure versus Protenix-Dock’s success in generating near-native poses. Left column: Ground truth structure with protein shown in green-cyan and ligand in light-grey. Middle column: Lowest-RMSD pose among top 50 poses sampled by Vina. Right column: Top-ranked pose from Protenix-Dock’s self-sampled and scored outputs. Heavy-atom RMSD values ( $\text{\AA}$ ) are indicated in parentheses. **A.** Angiotensin-1 converting enzyme N-domain in complex with dual ACE/NEP inhibitor AD012. **B.** Human galectin-3 carbohydrate recognition domain bound to a selective antagonist.